MONTHLY
WEATHER
REVIEW

VOLUME 94, NUMBER 8

AUGUST 1966

QUALITY CONTROL FOR PROBABILITY FORECASTS [1,2]

EDWARD S. EPSTEIN

Department of Meteorology and Oceanography, The University of Michigan, Ann Arbor, Mich.

ABSTRACT

The meaning of probabilistic weather forecasts is discussed from the point of view of a subjectivist concept of probability. The prior degree of belief of probabilities of the weather in question, for a given forecast statement, is expressed analytically as a beta function. Bayes' theorem is used to modify this degree of belief in the light of experience, producing a posterior degree of belief which is also in the form of a beta function. By establishing an arbitary criterion that one should always be able to assign at least as much belief to the probability interval implied by the forecast as to any other equivalent interval, a method of quality control for probability forecasts is developed. Appropriate tables are given to permit application of the method, and the implications of the method, for both forecaster and forecast user, are discussed.

1. INTRODUCTION

Increasingly, weather forecasts are being stated in terms of probabilities. There are numerous reasons for such a trend, but primarily this is being done because such probability statements are more useful than the conventional forecast to a broad segment of the users of the forecasts. The evidence for this is so clear (e.g., Thompson [8], [9]; Thompson and Brier [10]) that the wide use of probability statements would certainly have occurred much sooner, were it not for uncertainties regarding the use and interpretation of these statements on the part of the public and especially on the part of the meteorologists who must issue them. It is not reasonable to expect the meteorologist to use a language for communicating with the public if the meaning of the language is not clear to him. It is the purpose of this paper to examine the probability statement as it is normally used in forecasting, and to discuss its meaning and interpretation. Further, we shall present a set of guidelines by which the forecaster can judge whether his forecasts are adhering to some minimum standards. We will also point out how the use of such a system of quality control should contribute to confidence in and correct interpretation of the forecasts on the part of the using public.

[1] This research was supported in part by the U.S. Department of Commerce, Environmental Science Services Administration, Weather Bureau under contract Cwb-10847.

[2] Publication No. 98 from the Department of Meteorology and Oceanography, The University of Michigan.

2. AN INTERPRETATION OF THE PROBABILITY STATEMENT

The conceptual basis of our consideration of this problem is the "subjectivist" view of probability (e.g., Savage [6]; Schlaifer [7]). It is not possible to expound, here, on the details of this approach, but we will merely point out that this view corresponds well with the manner in which the term "probability" is used in everyday conversation. The emphasis, in this definition of probability, is on how it is used, and not, as in more classical approaches, on how it was derived. A number, between zero and one, can be a probability, if an individual is willing to assign that number as the "degree of belief" (Lindley [4]) he holds that the event will occur, and to act accordingly. There is no need to consider, even conceptually, a lengthly series of trials to determine the relative frequency of the event. This view of probability is especially useful and meaningful when applied to problems of decision-making (e.g., Epstein [2]). Thus it is an appropriate view to take in discussing meteorological probabilities, which we provide, after all, to help others make decisions concerning weather-sensitive activities. Also, the axiomatic treatment of probabilities which has appeared in probability texts (e.g., Feller [3]) for many years, and the resulting theorems and formulae, are equally applicable to probabilities having a "personal" or "subjectivist" basis.

The term "subjectivist" as used here should not be confused with the term "subjective" as used in contradistinction to "objective" and also applied to meteorological probabilities. The latter term refers to the manner in which the probabilities are *derived;* the former to the manner in which they are *used* or *interpreted.* From a subjectivist point of view, a probability must be *believed* to be meaningful. In general, I am far more likely to believe the probability suggested by a well-tested objective system than one based on an ill-defined subjective evaluation. On this basis my own personal probability (or the reader's, I suspect) will in general agree more readily with objectively, rather than subjectively, derived forecasts. In the same vein, the meteorologist issuing a forecast should *believe* his statement; it should be his personal probability whether or not it is wholly subjectively or objectively derived. Indeed, this is a good argument for the normal procedure of providing the forecaster with objective probabilities as guidance material, and allowing him to issue modified statements according to his further subjective judgment and belief.

For example, a forecaster may state that the probability of rain tomorrow is 0.3. The number, 0.3, should measure his "degree of belief" in the event, rain tomorrow. To the extent that the statement corresponds to the forecaster's belief, the forecaster implies that he would be indifferent between an outright gift of $3 or the chance to get $10 if and only if it rains tomorrow.[3] If, for the forecaster, the probability of rain tomorrow were greater than 0.3, he would choose the second option and vice versa. One could test whether or not a forecaster *believes* what he says by offering him such options and observing his actions. If there is a discrepancy, then the forecaster is not being honest and the stated number is not the probability, to him, of rain tomorrow, and may not be such a probability to anyone at all. However, this is not proper behavior for a forecaster. We must and shall expect our forecasters to be honest, both to themselves and to the public. Thus the number used by the forecaster is assumed to be *his* probability and the only evidence we require is his statement to that effect.

On the other hand an individual, say an ice cream vendor, may feel a twitching in his elbow which he usually accepts as a certain indicator of rain. Perhaps he should not even open his stand in the zoo tomorrow. Upon hearing the forecast, though, he decides to order a small supply of ice cream. In so doing, he is acting as though there were a relatively large probability of rain. For the sake of argument let us assume that his actions are those he would take if the probability of rain were 2/3.[4] So long as he acts in this way, then the probability of rain tomorrow *is*, for him, 2/3.

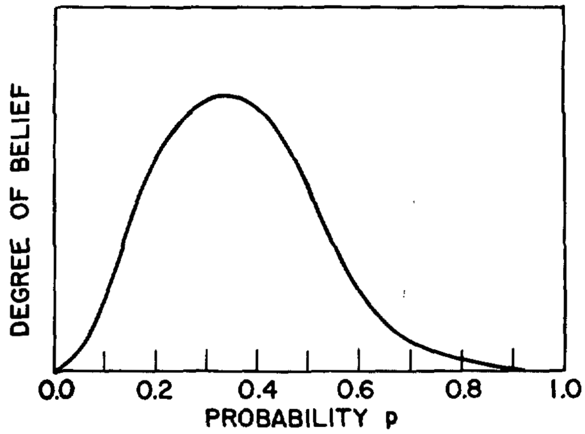

FIGURE 1.—One possible graphical statement of "degree of belief" when the forecaster's statement is $p=0.3$.

This example is not meant to imply that the user's meteorological judgment is better than, or even as good as, the forecaster's. But the user can and does make such judgments, and the probabilities he uses (whether or not they are explicitly stated) in making his decisions are as valid (in the sense of the meaning of probability, but likely not in the sense of verification) as those stated by the forecaster. The example also illustrates that the forecaster's stated probability will not, and indeed need not, be always taken at its face value by the user.

To illustrate this point further, let me examine what the forecaster's statement might mean to me. I assume here, for the purposes of argument, that this is a subjectively derived forecast, made by a trained and experienced forecaster, but that I have no knowledge of the forecaster's previous forecasting record or previous verification scores. Hearing the statement, 0.3, I am not at all confident that the probability of rain is exactly 0.3; I would not be terribly surprised if the probability were 0.2, or perhaps even 0.5. On the other hand, I would state, almost with certainty, that the probability of rain tomorrow is not 9/10. In other words, having heard the forecast, I have a mental image of my own "degree of belief" in what the probability of rain really is. This image might be represented by the curve in figure 1.

The more confidence I choose to place in the forecaster, the sharper the peak in my curve will be, and the more my belief will be concentrated near the stated probability. Indeed, if I knew the forecast had been based on an objective scheme in which I had confidence, my curve might be considerably more peaked. The area under the curve between any two values of the abscissa represents the extent to which I believe that the probability of rain falls between those two limits. Since I am certain that some probability between zero and one is the right one, I should place a scale on my degree of belief such that the area under the curve equals unity. Also a negative degree of belief would have no meaning to me, so I draw the curve so that it is everywhere either on or above the horizontal axis.

---

[3] Put in other terms, the forecaster is saying, "the odds are 3 to 7. Place your money and take your pick." We are assuming here that the utility of cash is linear over the amounts involved (cf. Schlaifer [7], pp. 41 ff.; Epstein [2]).

[4] The quantification is ours, and we introduce it for the sake of clarity of exposition; the businessman must quantify his order but rarely will he quantify his reason for that order.

A curve, or function, which has these properties is a probability density function. I will refer to it as the prior density of the probability of rain p, when the stated or forecast probability of rain is ρ. It is especially helpful to express the prior density in some convenient analytic form, $f(p)$. Since there is no a priori reason to expect the forecaster's judgment to be biased, it is reasonable to choose a formula such that the mean (or expected) value $(\int_0^1 pf(p)dp)$ of the probability of rain implied by the function is equal to ρ, the stated probability. I also include an adjustable parameter (L) which allows me to control the level of confidence I am willing to place in the forecaster. It is my hope that any reader, by choosing an appropriate value of L, will be able to use this particular functional form to give a prior density curve which corresponds sufficiently to his relative degree of belief.

3. ANALYTIC FORM FOR THE PRIOR DEGREE OF BELIEF

The analytic form that I choose for the prior density of p, the probability of rain for a particular forecast probability, ρ, is

$$f(p|\rho)=\frac{\Gamma(L)}{\Gamma(L\rho)\Gamma(L-L\rho)}\,p^{L\rho-1}(1-p)^{L(1-\rho)-1}, \qquad 0\leq p\leq 1 \qquad (1)^{5}$$

defined only for $0<\rho<1$ and $L>0$. In figure 2 are plots of several representative prior distributions for different values of L and ρ. Note the wide variety of shapes the density can take for different parameter values. I might point out that my own prior densities seem to correspond best to $L=10$.

Equation (1) is not a valid probability density for $\rho=0$ or $\rho=1$, since then its integral between zero and one does not converge. For these cases the following functional forms are suggested:

$$f(p|\rho=0)=\frac{\Gamma(L)}{\Gamma(L\epsilon)\Gamma(L-L\epsilon)}\,p^{L\epsilon-1}(1-p)^{L(1-\epsilon)-1}, \qquad 0\leq p\leq 1 \qquad (2)$$

$$f(p|\rho=1)=\frac{\Gamma(L)}{\Gamma(L\epsilon)\Gamma(L-L\epsilon)}\,p^{L(1-\epsilon)-1}(1-p)^{L\epsilon-1}, \qquad 0\leq p\leq 1 \qquad (3)$$

where ϵ is a small positive number.

Equations (1)–(3) are examples of beta density functions:

$$\beta(p|a,b)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\,p^{a-1}(1-p)^{b-1}.$$

The mean value of p is $a/(a+b)$ and its variance is $ab/[(a+b)^2(a+b+1)]$. Thus the mean of (1) is ρ and the variance is $\rho(1-\rho)/(L+1)$. For (2) and (3) the means are ϵ and $1-\epsilon$, respectively, and the variances are both $\epsilon(1-\epsilon)/(L+1)\simeq\epsilon/(L+1)$.

[5] The vertical bar within parentheses is standard notation for *conditional probabilities;* e.g., $f(p|\rho)$ means the density of p for a given value of ρ. The terms $\Gamma(X)$ represent gamma functions. When X is a positive integer $\Gamma(X)=(X-1)!$ and generally $\Gamma(X)=(X-1)\Gamma(X-1)$. When X is not an integer the value of the function may be taken from a table of the gamma function.

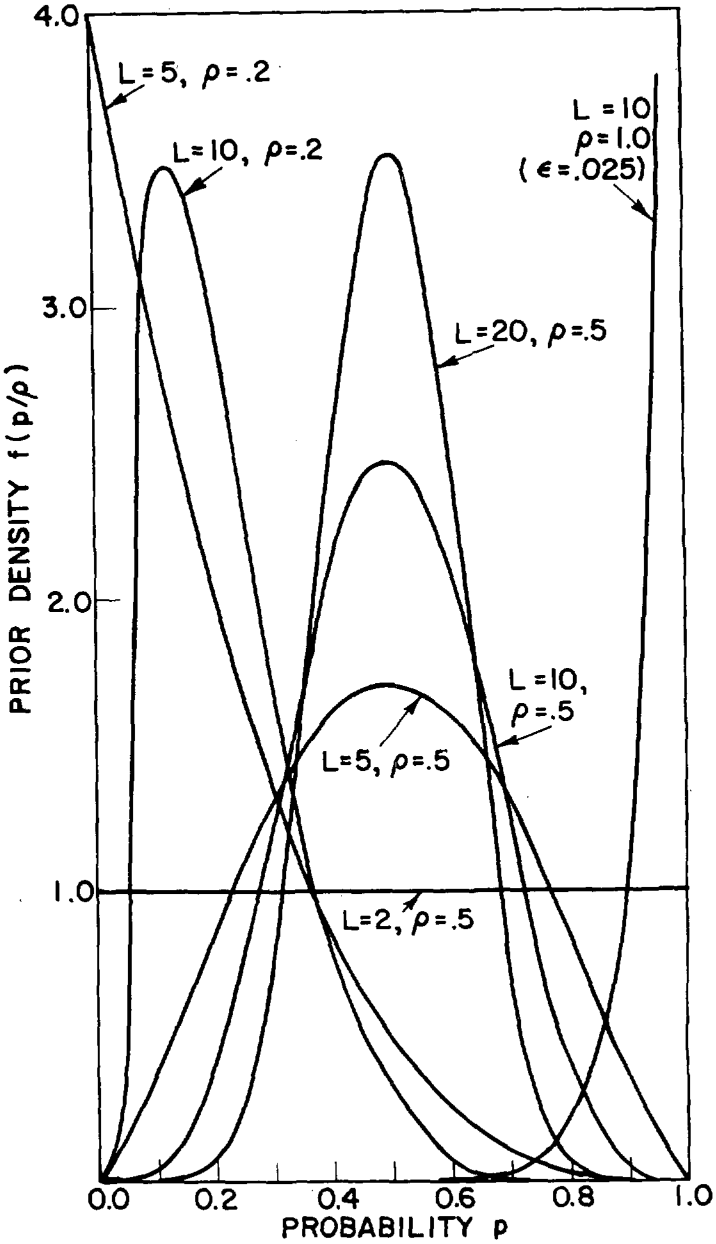

FIGURE 2.—Analytic form of prior degree of belief for several pairs of values of the parameters L and ρ.

4. REVISION OF THE PRIOR DEGREE OF BELIEF IN THE LIGHT OF EXPERIENCE

A standard forecast procedure is to permit the forecaster to make statements only in the form $\rho=0.0, 0.1, 0.2, \ldots, 0.9, 1.0$; i.e., he must state the probability of the weather in tenths. Once one has observed the forecaster's performance over some period of time it is reasonable that he should use that experience to revise his judgment as to what the statements mean to him. For example, if a forecaster has made the statement $\rho=0.3$ with regard to rain on ten occasions, and rain subsequently occurred six times, I would then consider it far more likely that the probability of rain, when this forecaster says $\rho=0.3$, is really 0.5, than that the probability of rain is 0.2; although previously, I might have thought that probabilities 0.2 and 0.5 were about equally likely.

The procedure which one follows in revising his estimate of the relative likelihood of the probabilities is based on a fundamental theorem of probability, known as Bayes' Theorem, which can be written

$$P(H|D)=\frac{P(H)P(D|H)}{P(D)}. \tag{4}$$

Simply stated, this theorem states that one's degree of belief that a particular hypothesis (H; say that $|p-\rho|<0.1$) is true, once some data (D) are available, $P(H|D)$, is proportional to the product of one's degree of belief in the hypothesis, $P(H)$, times the probability that the particular set of data would have occurred if the hypothesis were true, $P(D|H)$. The term $P(D)$ is a proportionality factor which is easy to evaluate and otherwise of little significance. For example, on the basis of the hypothesis (H_1) that the probability of rain is 0.2, the probability of six occurrences of rain in ten times (the data, D) is, from the binomial distribution,

$$P(D|H_1)=\frac{10!}{4!6!}(0.2)^6(0.8)^4=0.0055.$$

If the hypothesis were H_2, that the probability of rain is 0.5, then the probability of six occurrences of rain in ten times would be

$$P(D|H_2)=\frac{10!}{4!6!}(0.5)^6(0.5)^4=0.2050.$$

If, for the sake of argument, my prior judgment had been that $p=0.2$ was twice as likely as $p=0.5$ (when the forecaster said $\rho=0.3$), i.e., $P(H_1)=2P(H_2)$, then my posterior judgment would be that $p=0.2$ is only

$$\frac{P(H_1|D)}{P(H_2|D)}=\frac{2\times 0.0055}{0.2050}=0.0540$$

times as likely as $p=0.5$.

It is worth pointing out here that psychologists have found that people, in general, can make quite good intuitive judgments with regard to simple probabilities. However, when faced with a problem requiring the revision of these probabilities in the light of subsequent information, the same people exhibit rather poor intuitive judgments (Edwards [1]). Consequently, it is well that one should approach in a formal way this problem of revising prior beliefs in the light of subsequent information.

5. ANALYTIC FORM FOR THE POSTERIOR DEGREE OF BELIEF

Once again, let $f(p|\rho)$ be the prior density. We wish to determine a corresponding function representing our degree of belief, $g(p|\rho,n,r)$ after the forecast probability ρ has been stated n times, and the particular event (say rain, or subminimal ceiling) has subsequently occurred r times. We shall rewrite Bayes' Theorem, equation (4), as

$$g(p|\rho,n,r)=\frac{f(p|\rho)\lambda(r;n,p)}{\int_0^1 f(p|\rho)\lambda(r;n,p)dp}, \tag{5}$$

where $\lambda(r;n,p)$ is identified with $P(D|H)$ in equation (4) and is the probability of r occurrences, in n trials, of an event whose probability of occurrence on a single trial is p. This is, of course, the binomial distribution

$$\lambda(r;n,p)=\frac{n!}{r!(n-r)!}p^r(1-p)^{n-r}. \tag{6}$$

We also identify $g(p|\rho,n,r)$ with $P(H|D)$, $f(p|\rho)$ with $P(H)$, the prior probability of the hypothesis; $P(D)$, the proportionality factor, becomes the denominator in equation (5). Substituting, in (5), both (1) and (6), gives

$$g(p|\rho,n,r)=\frac{\Gamma(L+n)}{\Gamma(L\rho+r)\Gamma(L-L\rho+n-r)}\times \times p^{L\rho+r-1}(1-p)^{L(1-\rho)+n-r-1}. \tag{7}$$

The posterior density, $g(p|\rho,n,r)$, like the prior, is a beta density, when the likelihood, $\lambda(r;n,p)$, is binomial. The analytic simplicity and symmetry of this result is one of the reasons for choosing the beta density as the form of the prior degree of belief.

The posterior mean value of p, $\int_0^1 pg(p|\rho,n,r)dp$, once n and r are specified, is

$$E(p|\rho,n,r)=\frac{L\rho+r}{L+n} =\frac{L}{L+n}\rho+\frac{n}{L+n}\frac{r}{n}, \tag{8}$$

a weighted average of the prior mean, ρ, and the observed relative frequency, r/n. On this basis it is possible to interpret L as the equivalent number of observations, with relative frequency ρ, of the event in question, with which one is willing to credit the forecaster, a priori. The posterior variance

$$\text{Var}(p|\rho,n,r)=\frac{E(p|\rho,n,r)[1-E(p|\rho,n,r)]}{L+n+1}$$

decreases with both L and n. As data accumulate there is a decreasing range of p over which one has a substantial degree of belief. Also, as n increases, the importance of the parameters of the prior density decreases; $E(p|\rho,n,r)$ approaches r/n, and Var $(p|\rho,n,r)$ approaches $[(r/n)(1-r/n)]/n$.

If the prior densities (2) and (3), rather than (1), are used in (5), the resulting posterior distributions are

$$g(p|\rho=0,n,r)=\frac{\Gamma(L+n)}{\Gamma(L\epsilon+r)\Gamma(L-L\epsilon+n-r)}\times \times p^{L\epsilon+r-1}(1-p)^{L(1-\epsilon)+n-r-1}$$

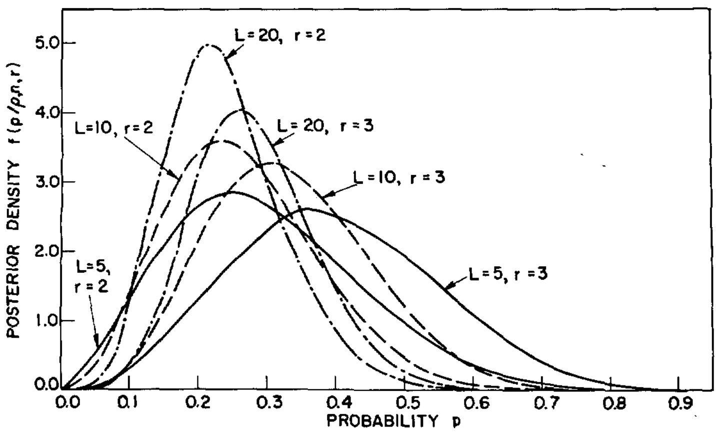

FIGURE 3.—Examples of posterior degree of belief for $\rho=0.2$ and $n=5$.

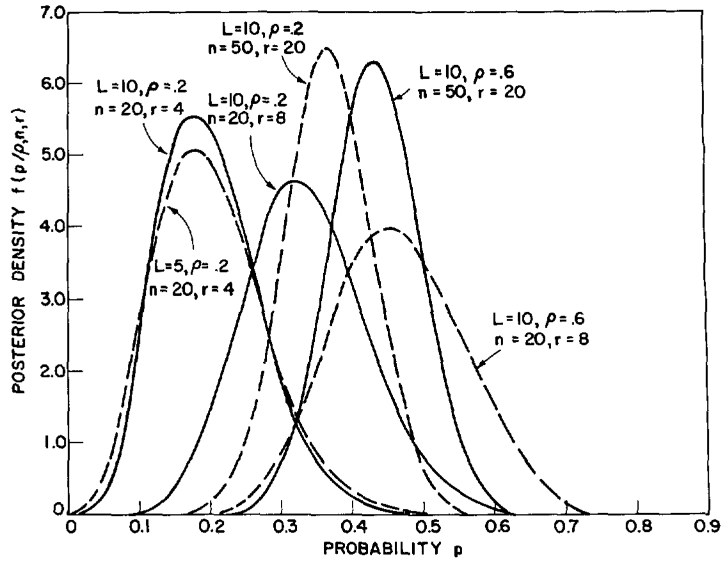

FIGURE 4.—Examples of posterior degree of belief.

$$g(p|\rho=1, n, r)=\frac{\Gamma(L+n)}{\Gamma(L\epsilon+r)\Gamma(L-L\epsilon+n-r)}\times$$
$$\times p^{L(1-\epsilon)+n-r-1}(1-p)^{L\epsilon+r-1}.$$

Some posterior distributions for various combinations of ρ, L, n, and r are shown in figures 3 and 4. One of the things to note is that for large n, not only does L become unimportant but also, in a sense, the value of ρ becomes unimportant. If, following 50 out of 100 times when a forecaster says $\rho=0.1$, it rains, then I will choose to interpret his statement, "$\rho=0.1$," as saying, for me, "the probability of rain is 0.5."

It may seem absurd that a forecaster would allow such a situation to arise. But we might have taken a less extreme example, in which case there would be considerable doubt as to whether the forecaster was biased in his estimations of the probabilities, or whether, by chance, an unusual series of events had occurred. In any case it would appear wise that the forecaster (and his supervisor) have some means to check on and eliminate such biases before they could possibly become so severe. The forecaster, as well as the forecast users, can reinterpret his statements. On this basis the forecaster should reinterpret first, revise his own belief, and, to remain honest, revise his statement. In this way the above situation should never arise. Furthermore, the user of the forecasts is not likely to maintain a complete set of records, and it would be most unfortunate if he chooses to believe $p=\rho=0.1$, when he should believe $p=0.5\neq\rho$. Also, if the user does keep records, we would like these to reflect favorably on the weather services.

6. QUALITY CONTROL FOR PROBABILITY FORECASTS

Before the forecaster began his predictions we had assigned a prior degree of belief to the various probabilities, depending upon what statement the forecaster would make. One way in which to look at the prior density is to consider the degree to which we believed, a priori, that the probability actually falls within the interval implied by the forecast statement; i.e., Prob $\{\rho-\delta/2<p<\rho+\delta/2\}$, where δ is the interval between standard values of the probability that are used in the forecast statements. From (1) we can compute

$$\text{Prob}\ \{\rho-\delta/2<p<\rho+\delta/2\}=\int_{\rho-\delta/2}^{\rho+\delta/2} f(p|\rho)dp.$$

Some results are shown in table 1. Similar results could be tabulated after a series of forecasts had been made. We show, for example, in table 2, the posterior likelihood that $|\rho-p|<0.05$, for $L=5$, $n=5$.

After five statements, "$\rho=0.3$," for example, we consider it more likely than previously that $0.25<p<0.35$, if there were either one or two occurrences of the pertinent weather. If the pertinent weather had occurred on all five occasions, our belief that $\rho=0.3$ means $0.25<p<0.35$ is only about 1/45. Indeed, for the particular case $L=5$, $\rho=0.3$, $n=5$, values of r of 4 or 5 make it appear so unlikely that p is in the interval implied by the forecast statement,

TABLE 1.—*Prior likelihood of* $\{\rho-0.05<p<\rho+0.05\}$

Forecast probability ρ	Confidence factor L: 5	10	20
0.0	0.862	0.841	0.840
0.1	.269	.399	.556
0.2	.206	.299	.421
0.3	.182	.263	.371
0.4	.172	.248	.349
0.5	.169	.243	.342
0.6	.172	.248	.349
0.7	.182	.263	.371
0.8	.206	.299	.421
0.9	.269	.399	.556
1.0	.862	.841	.840

TABLE 2.—*Posterior likelihood of* $\{\rho-0.05<p<\rho+0.05\}$ *for* $L=5$, $n=5$

Number of occurrences of pertinent weather r	Forecast probability ρ										
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.924	0.246	0.157	0.113	0.083	0.059	0.039	0.022	0.009	0.002	0.000
1	.310	.409	.299	.235	.188	.148	.112	.077	.043	.014	.000
2	.056	.223	.258	.259	.248	.230	.207	.175	.132	.070	.006
3	.006	.070	.132	.175	.207	.230	.248	.259	.258	.223	.056
4	.000	.014	.043	.077	.112	.148	.188	.235	.299	.409	.310
5	.000	.002	.009	.022	.039	.059	.083	.113	.157	.246	.924

TABLE 3.—*Limiting permissible range of r*

Forecast Probability ρ	Number of Forecasts n						
	5	10	15	20	25	50	
				$L=5$			
0.0	0–2	0–2	0–3	0–3	0–3	0–5	
0.1	0–2	0–3	0–4	0–5	1–6	1–11	
0.2	0–3	1–5	1–6	2–8	2–9	5–17	
0.3	0–3	1–6	2–8	3–10	4–12	9–22	
0.4	1–4	2–6	3–9	5–12	6–14	13–26	
0.5	1–4	3–7	5–10	6–14	8–17	18–32	$\delta=0.1$
0.6	1–4	4–8	6–12	8–15	11–19	24–37	
0.7	2–5	4–9	7–13	10–17	13–21	28–41	
0.8	2–5	5–9	9–14	12–18	16–23	33–45	
0.9	3–5	7–10	11–15	15–20	19–24	39–49	
1.0	3–5	8–10	12–15	17–20	22–25	45–50	
0.0	0–2	0–3	0–3	0–4	0–5	0–8	
0.2	0–3	0–5	1–6	1–8	1–10	4–18	
0.4	1–4	2–7	2–10	4–12	6–15	12–28	
0.6	1–4	3–8	5–13	8–16	10–19	22–38	$\delta=0.2$
0.8	2–5	5–10	9–14	12–19	15–24	32–46	
1.0	3–5	7–10	12–15	16–20	20–25	42–50	
				$L=10$			
0.0	0–2	0–2	0–3	0–3	0–4	0–6	
0.1	0–3	0–4	0–5	0–6	0–7	1–11	
0.2	0–4	0–5	1–7	1–8	2–10	4–17	
0.3	0–4	1–6	1–8	2–10	3–12	9–22	
0.4	0–4	1–7	3–10	4–12	5–15	13–27	
0.5	0–5	2–8	4–11	6–14	8–17	17–33	$\delta=0.1$
0.6	1–5	3–9	5–12	8–16	10–20	23–37	
0.7	1–5	4–9	7–14	10–18	13–22	28–41	
0.8	1–5	5–10	8–14	12–19	15–23	33–46	
0.9	2–5	6–10	10–15	14–20	18–25	39–49	
1.0	3–5	3–10	12–15	17–20	21–25	44–50	
0.0	0–2	0–3	0–4	0–5	0–5	0–8	
0.2	0–4	0–5	0–7	1–9	1–10	3–19	
0.4	0–5	1–7	2–10	4–13	5–15	12–29	
0.6	0–5	3–9	5–13	7–16	10–20	21–38	$\delta=0.2$
0.8	1–5	5–10	8–15	11–18	15–24	31–47	
1.0	3–5	7–10	11–15	15–20	20–25	42–50	
				$L=20$			
0.0	0–4	0–3	0–3	0–4	0–4	0–6	
0.1	0–4	0–5	0–6	0–7	0–8	0–12	
0.2	0–5	0–6	0–8	0–9	1–11	4–18	
0.3	0–5	0–7	0–9	1–11	2–13	8–23	
0.4	0–5	0–8	1–11	3–13	4–16	12–28	
0.5	0–5	1–9	3–12	5–15	7–18	17–33	$\delta=0.1$
0.6	0–5	2–10	4–14	7–17	9–21	22–38	
0.7	0–5	3–10	6–15	9–19	12–23	27–42	
0.8	0–5	4–10	7–15	11–20	14–24	32–46	
0.9	1–5	5–10	9–15	13–20	17–25	38–50	
1.0	1–5	7–10	12–15	16–20	21–25	44–50	
0.0	0–3	0–4	0–5	0–5	0–6	0–9	
0.2	0–5	0–7	0–8	0–10	0–12	2–20	
0.4	0–5	0–9	1–11	2–14	4–17	10–30	$\delta=0.2$
0.6	0–5	1–10	4–14	6–18	8–21	20–40	
0.8	0–5	3–10	7–15	10–20	13–25	30–48	
1.0	2–5	6–10	10–15	15–20	19–25	41–50	

that we can tell at a glance that there must be another interval of p, of the same size, which is more likely, a posteriori. In general, a statement of this sort can be made whenever

$$\text{Prob}\{\rho-\delta/2<p<\rho+\delta/2|\rho, L, n\}<\delta \qquad (9)$$

which is equivalent to saying that experience suggests strongly that the forecaster's probability statements are biased. We have used this somewhat arbitrary criterion to define a "limiting permissible range of r", for various values of n, ρ, and L. Results are given in table 3. Values of r falling outside these limits can be taken as evidence that the methods of estimating the probabilities are biased and require some adjustment. One can look upon table 3 as forming the basis of a technique for quality control of probability forecasts.

It seems unlikely that the limiting values in this table will often be reached. A forecaster will (and should) generally keep a tally on his own performance. Consider a forecaster for whom $L=5$. If, for example, after four forecasts in which he has stated $\rho=0.3$, he finds no observed cases of the pertinent weather, he will be concerned over the possibility of bias, but not overly so, since he knows that (a) zero successes out of four trials when the probability is actually 0.3 is not a very unlikely event (probability equals 0.240), and (b) even if the weather does not occur the next time he states 0.3, he will still lie within the indicated permissible range. On the other hand, if the same results ($n=4$, $r=0$) had occurred for statements $\rho=0.4$, the same forecaster would, and should, be more concerned. Indeed, it is considerably less likely that no successes would occur after four trials with $p=0.4$ (probability equals 0.130). Very likely, the next time he states "$\rho=0.4$", he will believe $\rho>0.4$ according to his original method of estimating, but he will be adjusting his method of estimation to avoid falling outside the lower limit, $r=1$, for $n=5$, $\rho=0.4$, $L=5$. Put another way, the forecaster is now revising his prior judgment on the basis of his experience. On the basis of the earlier forecasts, though, it is reasonable that the forecaster should now consider his earlier estimates of $\rho=0.4$ as overestimates of the probability, and he should include in that category situations with greater likelihood of the particular weather event in question. Note, also, that if the forecaster had greater self-confidence, he would have had less urge to modify his procedure. Indeed for $L=10$, $n=5$, $\rho=0.4$, the value $r=0$ is within the permitted range.

This table, then, serves best as a guide to the forecaster to help him keep tab on his probability statements, and to keep him from issuing biased statements. Again, as n gets large, the value of L loses significance, and the forecaster's subsequent experience outweighs any prior judgment concerning his abilities.

7. IMPLICATIONS OF QUALITY CONTROL FOR THE FORECAST USER

From the user's point of view, the most important single aspect of a probability forecast is the expected value, $E(p)$. He will normally use the forecast statement as though it were the expected value. In other words he will act as though $E(p|\rho, n, r) = \rho$, even though he has no knowledge of n and r. Indeed, he will in effect be using only his prior distribution, for which $E(p|\rho) = \rho$. It is instructive to examine the ranges of $E(p|\rho, n, r)$ which might result if the forecaster manages to adhere to the limits given in table 3. These values are easily derived from table 3 by applying equation (8). Some results are shown in table 4.

TABLE 4.—*Range of* $E(p|\rho, n, r)$

Forecast Probability ρ	$L=5$			$L=10$		
	Number of Forecasts n			Number of Forecasts n		
	5	20	50	5	20	50
0.0	0.01–0.20	0.00–0.12	0.00–0.09	0.01–0.15	0.00–0.13	0.00–0.10
0.1	0.05–0.25	0.02–0.22	0.03–0.22	0.07–0.27	0.03–0.23	0.03–0.20
0.2	0.10–0.40	0.12–0.36	0.11–0.33	0.13–0.40	0.10–0.33	0.10–0.32
0.3	0.15–0.55	0.18–0.42	0.19–0.43	0.20–0.47	0.17–0.43	0.20–0.42
0.4	0.30–0.60	0.28–0.56	0.27–0.51	0.27–0.53	0.27–0.53	0.28–0.52
0.5	0.35–0.65	0.34–0.62	0.37–0.63	0.33–0.67	0.37–0.63	0.37–0.63
0.6	0.40–0.70	0.44–0.72	0.49–0.73	0.47–0.73	0.47–0.73	0.48–0.72
0.7	0.55–0.85	0.54–0.82	0.57–0.81	0.53–0.80	0.57–0.83	0.58–0.80
0.8	0.60–0.90	0.64–0.88	0.67–0.89	0.60–0.87	0.67–0.90	0.68–0.90
0.9	0.75–0.95	0.78–0.98	0.78–0.97	0.73–0.93	0.77–0.97	0.80–0.97
1.0	0.80–0.99	0.88–1.00	0.91–1.00	0.85–0.99	0.87–1.00	0.90–1.00

For very large n, the values in table 4 should approach $\rho \pm 0.05$, but we can see by examining the actual results that this limit is approached quite slowly. Nevertheless, even for small values of n, the posterior expected value of p never can differ by more than about 0.2 from ρ, if the forecaster stays within the limits of table 3. For $n \geq 20$, $|\rho - E(p|\rho, n, r)| < 0.13$. This means, to the user of the forecast, that he need not keep a record on the forecaster's performance, for the expected value of p that he would believe, if he had all the information contained in that record, will never be very different from the value he believes (ρ) by simply accepting the forecaster's statement. In other words, this method of quality control acts as an assurance to the customers that they will not go very far wrong by accepting the forecasts at their face value.

8. CONCLUSIONS

Probability statements in weather forecasts are slowly becoming the rule instead of the exception. In spite of their clear superiority over categorical statements as a tool in the decision-making process, lack of understanding, largely on the part of meteorologists, and lack of confidence in their acceptance by the public, have impeded their general introduction. In the present paper I have attempted to overcome, at least in part, some of the meteorologists' reticence to use the language of probability.

On the basis of a logical and self-consistent interpretation of the meaning of the probability statement, a guide for the control of the quality of these statements has been developed. The forecaster, by use of this guide, can adjust his own judgments in the light of his personal experience, and at the same time assure the public that his probability statements can be taken at their face value; i.e., that they are essentially unbiased.

In the course of our arguments we have had to introduce a certain degree of arbitrariness. Perhaps most serious in this regard is the use of a particular form for the prior degree of belief. There is no possible way of devising a universally acceptable prior density, and it is possible that many readers hold strongly divergent views. It is my opinion, however, that the prior density employed can be accepted as *reasonable* by most readers. Furthermore I am confident that if one were to introduce any other *reasonable* prior distribution the results would be essentially unchanged (although the mathematics might become considerably more difficult).

Another point that I would emphasize is that the minimum standards for probability statements given in table 3 do not guarantee an excellent score on any particular verification system. Certainly the forecaster who uses these controls will almost certainly have a probability score[6] of less than 0.5; i.e., the forecasts cannot be very bad. But forecasters can, in general, do considerably better (Sanders [5]), which points up the suggestion made earlier that in actual practice the forecaster, left to his own devices, would normally stay well within the control limits.

ACKNOWLEDGMENTS

This paper has benefited from the comments and suggestions of Mr. Christopher Hayden, and Mr. Allen H. Murphy. Mr. Murphy in particular has helped in a number of significant revisions and has overseen the typing and the drawing of the figures. The computations required to develop the various tables turned out to be an extremely difficult task, and the author is most grateful for the efforts of Mr. David Childs, who wrote a program for evaluating the incomplete beta function ratio. To the best of our knowledge this had not been previously accomplished for arbitrary parameters which could take on large values. Finally, I would like to express my appreciation to the Weather Bureau, ESSA, for their support for this research, and in particular Dr. William Klein and Mr. Roger Allen of the Techniques Development Laboratory for their advice and comments.

REFERENCES

1. W. Edwards, "Subjective Probabilities Inferred from Decision," *Psychological Review*, vol. 69, No. 2, Mar. 1962, pp. 109–135.
2. E. S. Epstein, "A Bayesian Approach to Decision Making in Applied Meteorology," *Journal of Applied Meteorology*, vol. 1, No. 2, June 1962, pp. 169–177.
3. W. Feller, *An Introduction to Probability Theory and Its Applications*, John Wiley and Sons, Inc., New York, 1950, 419 pp.

[6] In the situation discussed here, the probability score is

$$\frac{2}{N}\sum_{i=1}^{n}(\rho_i - O_i)^2$$

where O_i equals one if the pertinent weather occurs, and zero otherwise, and $i=1, 2, \ldots, N$ distinguishes among N individual forecast statements.

4. D. V. Lindley, *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part I, Probability,* Cambridge University Press, London, 1965, 259 pp.
5. F. Sanders, "On Subjective Probability Forecasting," *Journal of Applied Meteorology,* vol. 2, No. 2, Apr. 1963, pp. 191–201.
6. L. J. Savage, *The Foundations of Statistics,* John Wiley and Sons, Inc., New York, 1954, 294 pp.
7. R. Schlaifer, *Probability and Statistics for Business Decisions,* McGraw-Hill Book Company, Inc., New York, 1959, 732 pp.
8. J. C. Thompson, "On the Operational Deficiencies in Categorical Weather Forecasts," *Bulletin of the American Meteorological Society,* vol. 33, No. 6, June 1952, pp. 223–226.
9. J. C. Thompson, "Weather Decision Making—The Payoff," *Bulletin of the American Meteorological Society,* vol. 44, No. 2, Feb. 1963, pp. 75–78.
10. J. C. Thompson and G. W. Brier, "The Economic Utility of Weather Forecasts," *Monthly Weather Review,* vol. 83, No. 11, Nov. 1955, pp. 249–254.

[*Received May 13, 1966*]